

Structured Presentation Plan

Slide 1: Title Slide

Title: Analyzing Pregnancy-Related Stress Using Reddit Data

Subtitle: Semantic and Sentiment Analysis with Visual Insights

Slide 2: Introduction

Title: Introduction

Content:

- Objective:** Understand how stress manifests in text during pregnancy.
- Motivation:** Reddit provides a rich dataset of real-world experiences, offering insights into pregnancy-related stress.
- Approach:** Use advanced text analysis, including correlation, semantic analysis, and sentiment analysis, to uncover patterns.

Slide 3: Data Labeling and Correlation Analysis

Title: Data Labeling and Correlation Analysis

Content (English):

- Process:** Tagged 100 posts with two metrics: Stress Intensity and Emotional Overload.
- Pearson Correlation:** Calculated correlation to assess the relationship between these metrics.
- Result:** Strong positive correlation ($r > 0.7$).
- Why This Graph:** This analysis was conducted to validate whether Stress Intensity and Emotional Overload are related enough to be combined into a single metric.
- Conclusion:** A new metric, "Average Stress and Emotional Overload," was created to combine the two measures.

Content (Hebrew):

- תייגנו 100 פוסטים עם שני מדדים: עוצמת לחץ ועומס רגשי: התהליך.
- חישבנו את הקורלציה כדי להעריך את הקשר בין המדדים: קורלציית פירסון.

- נמצאה קורלציה חיובית חזקה: **תוצאה** ($r > 0.7$).
- ניתוח זה בוצע כדי לבדוק אם יש קשר מספיק חזק בין עוצמת: **למה הגרף הזה** הלחץ לעומס הרגשי כדי לשלבם למדד אחד.
- יצרנו מדד חדש, "ממוצע עוצמת לחץ ועומס רגשי", שמשלב את שני: **מסקנה** המדדים.

Graph: Emotional Overload vs Stress Correlation.

Slide 4: Pre and Post Analysis

Title: Stress Levels Before and During Protests

Content (English):

- **Objective:** Compare stress levels before and during protests (pre-BLM vs post-BLM).
- **Why This Test:** A T-Test was performed to determine whether there is a statistically significant difference in stress levels between the two periods.
- **Findings:** No significant difference was observed ($p > 0.5$).
- **Conclusion:** The lack of difference suggests that the protests did not significantly impact reported stress levels in the analyzed posts. This variable was excluded from further analysis.

Content (Hebrew):

- **מטרה** (BLM-ואחרי BLM-לפני) להשוות את רמות הלחץ לפני ובמהלך המחאות.
- כדי לבדוק האם יש הבדל מובהק סטטיסטית ברמות T בוצע מבחן: **למה המבחן** הלחץ בין שתי התקופות.
- **ממצאים** ($p > 0.5$) לא נמצא הבדל מובהק.
- היעדר ההבדל מצביע על כך שהמחאות לא השפיעו באופן משמעותי על: **מסקנה** רמות הלחץ המדווחות בפוסטים שניתחנו. משתנה זה לא נכלל בניתוחים נוספים.

Graph: Density Plot of Stress Levels by Period.

Slide 5: Semantic Analysis - TF-IDF

Title: Key Terms Identified with TF-IDF

Content (English):

- **Method:** TF-IDF was applied to identify the most significant terms in the dataset.
- **Findings:**
 - Words like "baby," "weeks," and "pregnancy" were most frequently discussed.
 - Emotional and mental health terms (e.g., "feel," "know") also stood out.
- **Why This Analysis:** This was conducted to highlight the core themes discussed in pregnancy-related posts.
- **Conclusion:** These terms highlight the primary themes of the discussions, focusing on medical and emotional concerns.

Content (Hebrew):

- כדי לזהות את המילים המשמעותיות ביותר בנתונים TF-IDF יישמנו: שיטה.
- **ממצאים:**
 - היו הנפוצות ביותר "pregnancy" ו-"weeks," "baby" מילים כמו.
 - בלטו גם כן ("know," "feel" כמו) מילים הקשורות לרגשות ובריאות נפשית.
- **למה הניתוח הזה:** ניתוח זה בוצע כדי להדגיש את הנושאים המרכזיים שנדונים: בפורסטים הקשורים להריון.
- **מסקנה:** המילים הללו מדגישות את הנושאים המרכזיים בדיונים, המתמקדים: בבריאות רפואית ורגשית.

Graph: TF-IDF Top Words.

Slide 6: Semantic Analysis - N-Grams

Title: Common Language Patterns: N-Grams

Content (English):

- **Method:** Extracted bigrams and trigrams to analyze repetitive language patterns.
- **Findings:**
 - Common bigrams: "I have," "in the hospital."
 - Common trigrams: "I feel like," "I have been."

- **Why This Analysis:** To uncover recurring linguistic patterns that highlight concerns or emotional states.
- **Conclusion:** These patterns reveal themes of emotional and medical challenges faced during pregnancy.

Content (Hebrew):

- חילצנו ביגרמים וטריגרמים כדי לנתח דפוסים שפה חוזרים: שיטה
- **ממצאים:**
 - ביגרמים נפוצים: "I have," "in the hospital."
 - טריגרמים נפוצים: "I feel like," "I have been."
- **למה הניתוח הזה:** כדי לחשוף דפוסים שפה חוזרים שמדגישים דאגות או מצבים רגשיים.
- **מסקנה:** דפוסים אלה חושפים נושאים של אתגרים רגשיים ורפואיים שנחוו במהלך ההריון.

Graph: Bigrams Analysis.

Slide 7: Text Length Analysis

Title: Text Length by Stress Levels

Content (English):

- **Objective:** Analyze the relationship between text length and stress levels.
- **Findings:**
 - Posts from individuals with higher stress levels tend to be longer.
 - Suggests that higher stress leads to more elaborate expressions.
- **Why This Analysis:** To determine whether text length can serve as an indirect measure of stress levels.
- **Conclusion:** Text length serves as a proxy indicator of stress levels.

Content (Hebrew):

- לנתח את הקשר בין אורך הטקסט לרמות הלחץ: מטרה
- **ממצאים:**
 - פוסטים מאנשים עם רמות לחץ גבוהות נוטים להיות ארוכים יותר.
 - מצביע על כך שלחץ גבוה מוביל לביטוי מפורט יותר.

- כדי לבדוק האם אורך הטקסט יכול לשמש כמדד עקיף לרמות לחץ.
- אורך הטקסט משמש כאינדיקטור עקיף לרמות לחץ: **מסקנה**.

Graph: Text Length by Stress Levels.

Slide 8: Sentiment Analysis (VADER)

Title: Sentiment Analysis Using VADER

Content (English):

- Objective:** Examine the emotional tone of posts using VADER sentiment analysis.
- Findings:**
 - Negative Sentiment:** Increases with stress levels, reflecting emotional distress.
 - Positive Sentiment:** Remains relatively stable across stress levels.
- Why This Analysis:** To quantify the emotional impact of stress and evaluate sentiment trends.
- Conclusion:** Negative sentiment correlates strongly with stress, while positive sentiment is less affected.

Content (Hebrew):

- VADER לבחון את הטון הרגשי של הפוסטים באמצעות ניתוח סנטימנט: **מטרה**.
- ממצאים:**
 - עולה עם רמות הלחץ, ומשקף מצוקה רגשית: **סנטימנט שלילי**.
 - נותר יחסית יציב בין רמות הלחץ: **סנטימנט חיובי**.
- כדי לכמת את ההשפעה הרגשית של הלחץ ולהעריך מגמות: **למה הניתוח הזה** סנטימנט.
- סנטימנט שלילי מקושר באופן חזק ללחץ, בעוד שסנטימנט חיובי מושפע: **מסקנה** פחות.

Graph: Negative Sentiment vs Stress.

Slide 9: Summary and Conclusions

Title: Summary and Future Directions**Content (English):**

- Key Findings:**

- High correlation between Stress Intensity and Emotional Overload.
- No significant difference in stress levels before and during protests.
- Negative sentiment correlates with higher stress, while text length increases with stress levels.
- Key terms and frequent phrases reflect emotional and medical concerns.

- Future Work:**

- Expand the dataset for better generalization.
- Use advanced machine learning techniques for deeper sentiment and topic analysis.

Content (Hebrew):

- ממצאים מרכזיים:**

- קורלציה גבוהה בין עוצמת לחץ לעומס רגשי.
- אין הבדל מובהק ברמות הלחץ לפני ובמהלך המחאות.
- סנטימנט שלילי מקושר לרמות לחץ גבוהות, בעוד שאורך הטקסט עולה עם רמות הלחץ.
- מונחים ותבניות נפוצים משקפים דאגות רגשיות ורפואיות.

- עבודה עתידית:**

- להרחיב את מאגר הנתונים לשם הכללה רחבה יותר.
- להשתמש בטכניקות מתקדמות של למידת מכונה לניתוח סנטימנט ונושאים מעמיק יותר.

Graph: None.

Detailed Graph Insights

- Emotional Overload vs Stress Correlation Graph:** Demonstrates a strong positive correlation, justifying the combination into one metric.

2. **Density Plot of Stress Levels by Period:** Shows no significant difference between stress levels pre- and post-protests.
3. **TF-IDF Top Words Graph:** Highlights key terms, including medical and emotional words, central to the dataset.
4. **Bigrams Analysis Graph:** Displays commonly used phrases, emphasizing health and emotional states.
5. **Text Length by Stress Levels Graph:** Correlates longer texts with higher stress levels, indicating a deeper need for expression.
6. **Negative Sentiment vs Stress Graph:** Clearly shows an increase in negative sentiment as stress levels rise, highlighting emotional impact.

סיכום תהליך עבודה מפורט

מבוא:

מטרה ורקע

- המחקר מבקש בניתוח את רמת הלחץ של נשים בהריון על פרסומים רפואיים ב Reddit.
- חקרנו כיצד רמת הלחץ באה לידי ביטוי בטקסטים על חוויות הריון, והאם ניתן להפיק תובנות משמעותיות מנתונים אלה.

תיג נתונים והשוואת קורלציית פירסון

למה חשבנו

- עומס רגשי-עוצמת לחץ ו תיגנו 100 פוסטים על פי שני מדדים.
- ביצענו קורלציית פירסון כדי לבדוק האם קיים קשר בין המדדים.

מה נמצאה בניתוח

- נמצאה קורלציה חזקה (עליה מ-0.7) שהצביעה על קשר משמעותי בין המדדים.
- ממוצע לחץ ועומס " בשל הקשר בין המדדים, נוצרה עמודה חדשה של: מסקנה "רגשי".

גרף:

- גרף המציג את הקורלציה בין המדדים. מסקנתנו: המדדים חופפים ומצדיקים יצירת משתנה משולב.

השוואה בין לפני-מחאות לאחר המחאות בארה"ב**T? למה עשינו מבחן**

- לבדוק האם רמת הלחץ היתה שונה לפני ואחרי המחאות: **מטרה**.
- ($p\text{-value} > 0.5$) לא נמצא הבדל מובהק: **תוצאה**.

מסקנות:

- המסקנה היתה שאין הבדל מובהק בין התקופות. משתנה זה לא שימש בניתוחים נוספים.

גרף:

- כולל תוצאות מבחן (density plot), גרף השוואה בין התקופות לפני ואחרי המחאות. המצורפות בגרף T.

(TF-IDF ו-N-grams בעזרת) ניתוח סמנטי**מטרות:**

- לזהות מילים ייחודיות ונפוצות בתוכן הפוסטים.
- לזהות דפוסי שפה נפוצים (ביגרמים וטריגרמים).

ממצאים:

- TF-IDF:** היו מרכזיות בדיון "baby", "weeks", "pregnancy" המילים.
- N-grams:** היו נפוצים, ומעידים על נושאים "I have a", "I feel like" ביטויים כמו רגשיים ורפואיים שחוזרים על עצמם.

מסקנות:

- ניתוח סמנטי זה מדגיש את התחומים שמעסיקים נשים בהריון.

גרפים:

- המציג את המילים המרכזיות בטקסטים TF-IDF גרף.
- גרף ביגרמים וטריגרמים המציג ביטויים נפוצים.
- "I" גרף זה מראה את הרצפים הדו-מילתיים הנפוצים ביותר, כמו: **ביגרמים** "was", "in the", "the", "of" אופי הטקסטים היומי-יומי, המצביעים על אופי הטקסטים היומי-יומי.

4. **גרף טריגרמים:** גרף זה מראה את הרצפים התלת-מילתיים, שמדגישים ביטויים: "I have a", "I feel like", משמעותיים יותר כמו שמצביעים על התבטאויות רגשיות, ורפואיות חוזרות.

ניתוח סנטימנט (VADER):

מטרות:

- לבדוק האם סנטימנט הטקסט משתנה לפי רמת הלחץ.
- לבחון את הקשר בין סנטימנט שלילי, חיובי ורמת הלחץ הממוצעת.

ממצאים:

- עולה עם עליית רמות הלחץ: **סנטימנט שלילי**.
- אינו משתנה בצורה מובהקת ביחס לרמת הלחץ: **סנטימנט חיובי**.

מסקנות:

- סנטימנט שלילי משקף את העלייה ברמת הלחץ, בעוד שסנטימנט חיובי נותר קבוע יחסית.

גרפים:

- גרף המציג את הקשר בין סנטימנט שלילי לרמת הלחץ הממוצעת (scatter plot).
- גרף המציג את ההתפלגות של סנטימנט חיובי ושלילי לפי רמות הלחץ (box plot).
- מדגיש את העלייה בסנטימנט: **גרף התפלגות סנטימנט שלילי לפי רמות הלחץ**. השלילי במקביל לעלייה ברמות הלחץ הגבוהות.
- מסביר שהשפעת הלחץ על סנטימנט חיובי פחות: **גרף סנטימנט חיובי מול לחץ**. משמעותית בהשוואה לסנטימנט שלילי.

ניתוח אורך הטקסט:

מטרות:

- לבדוק האם יש קשר בין רמות לחץ לאורך הטקסטים.

ממצאים:

- פוסטים בעלי רמת לחץ גבוהה נוטים להיות ארוכים יותר.

מסקנות:

- אורך הטקסט עשוי לשמש כסמן עקיף לעוצמת הלחץ שחווים הכותבים.

גרף:

- (low, medium, high) המציג את אורכי הטקסטים לפי רמות הלחץ box plot גרף

ניתוח התפלגות טקסט לפי שימוש בסימני פיסוק ואימוג'ים

מטרות:

- לבחון את דפוסי השימוש בסימני פיסוק ואימוג'ים בהקשר של לחץ רגשי

ממצאים:

- פוסטים עם לחץ גבוה מכילים יותר סימני שאלה וסימני קריאה
- שימוש באימוג'ים היה נדיר יחסית

מסקנות:

- דפוסי הפיסוק עשויים להצביע על רמת הלחץ או המתח הרגשי של הכותבים

גרף:

- (bar chart) גרף המציג את השימוש בסימני פיסוק ואימוג'ים

סיכום ותובנות

תובנות מרכזיות

- יש קשר משמעותי בין מדדי לחץ שונים ובין תוכן הטקסטים
- ניתוח סנטימנט ואורך הטקסט חושפים תבניות מעניינות המשקפות את חוויות המשתמשים

הצעות להמשך

1. ניתוח נרחב יותר על קבוצות דמוגרפיות שונות
2. בחינת גורמים נוספים שמשפיעים על רמות הלחץ בתוכן

שאלת המחקר

?באיזו מידה שבוע ההיריון משפיע על רמות הלחץ

התייחסות לשאלת המחקר

שאלת המחקר המרכזית אכן מעלה תהיות על הקשר בין שבוע ההיריון לבין רמות הלחץ. בניתוח שנעשה

1. שלבים שנעשו בהתייחסות לשאלת המחקר

- נבדק הקשר בין מדדים שונים של לחץ (עומס רגשי ועוצמת לחץ)

- ניתוח פוסטים כדי להבין דפוסים בטקסטים הקשורים ללחץ.
- על רמות הלחץ כמייצג אפשרי (pre/post) בדיקת השפעת תקופות להשפעות חיצוניות, אך נמצא כי ההשפעה אינה משמעותית.
- הוספת עמודת ממוצע בין עומס רגשי לעוצמת לחץ כדי לפשט את ההתייחסות למשתנים.

2. מסקנות שמסבירות את מיקוד המחקר:

- הניתוח הסמנטי והסנטימנטלי מדגיש את נושאי השיח ואת הקשר שלהם לרמות הלחץ.
- למרות שלא נבחנו באופן ישיר ההשפעות של שבוע ההיריון, דפוסי הטקסטים יכולים להצביע על מאפיינים עונתיים או הקשרים לשבועות מסוימים.

3. האם עמדנו בשאלת המחקר?

- במובן מסוים, כן. המחקר הצביע על גורמים כלליים שמשפיעים על רמות הלחץ (כמו נושאי השיח, סנטימנטים ואורכי טקסטים).
- עם זאת, לא נעשתה הפרדה ברורה שמסווגת לחץ לפי שבועות ההיריון, דבר שיכול להיות מיקוד למחקר עתידי.

Questions and Detailed Answers in English and Hebrew:

What is the reliability of the data?

English: The data reliability was ensured by a robust labeling process of 100 posts, focusing on stress-related metrics like emotional overload and stress intensity. These labels were consistent and validated through Pearson correlation, showing strong interdependence between the metrics. While the data provides meaningful insights, its generalizability may be limited due to the specific Reddit community and sample size.

Hebrew: אמינות הנתונים הובטחה על ידי תהליך תיוג מדויק של 100 פוסטים, תוך התמקדות במדדי לחץ כמו עומס רגשי ועוצמת לחץ. תיוגים אלו הראו עקביות ואומתו באמצעות קורלציית פירסון, שהצביעה על קשר חזק בין המדדים. עם זאת, ייתכן כי הכללת וגודל המדגם Reddit הממצאים מוגבלת בשל מאפייני קהילת

The results of your labeling (does it tell you anything interesting already?).

English: Labeling results indicated a strong correlation between emotional overload and stress intensity (Pearson correlation > 0.7). This insight justified creating a new variable that averaged these two metrics, simplifying further analyses. Additionally, labeled data revealed patterns such as longer posts being linked to higher stress levels.

Hebrew: תוצאות התיוג הצביעו על קשר חזק בין עומס רגשי לעוצמת לחץ (קורלציית ממצא זה הצדיק יצירת משתנה חדש הממוצע בין שני מדדים אלו, מה שהפך > 0.7 פירסון את הניתוחים הבאים לפשוטים יותר. בנוסף, הנתונים המתוייגים חשפו דפוסים כמו פוסטים ארוכים יותר שמקושרים לרמות לחץ גבוהות יותר.

What methods you used and what statistics you computed?

English: We used several methods:

1. **Pearson Correlation** to assess relationships between emotional overload and stress intensity.
2. **T-tests** to compare stress levels before and during protests, examining statistical significance.
3. **Semantic Analysis** through TF-IDF and N-grams to identify key words and patterns in the text.
4. **Sentiment Analysis (VADER)** to evaluate positive and negative sentiments based on stress levels.
5. Descriptive statistics and visualizations like scatter plots, box plots, and bar charts to interpret patterns.

Hebrew: השתמשנו במספר שיטות:

1. **קורלציית פירסון** לבדיקת קשרים בין עומס רגשי לעוצמת לחץ.
 2. **לשוואת רמות הלחץ לפני ובמהלך מחאות**, תוך בדיקת מובהקות **T מבחני** סטטיסטית.
 3. **לזיהוי מילות מפתח ודפוסים** בטקסט TF-IDF ו-N-grams **ניתוח סמנטי** באמצעות.
 4. **להערכת סנטימנטים חיוביים ושיליים בהתאם לרמות (VADER) ניתוח סנטימנט** הלחץ.
 5. **scatter plots, box plots ו-bar charts** סטטיסטיקה תיאורית וויזואליזציות כמו **לפרשנות דפוסים**.
-

Why did you choose this representation? Why these statistics?

English: These methods were chosen to address specific research objectives. Pearson correlation quantified relationships between stress metrics, while T-tests tested significant differences across periods. TF-IDF and N-grams captured semantic richness, and VADER provided sentiment polarity insights. Visualization techniques offered intuitive interpretations, making complex data accessible.

Hebrew: שיטות אלו נבחרו כדי לענות על מטרות המחקר. קורלציית פירסון כימתה את Pearson correlation, בדקו הבדלים משמעותיים בין תקופות T הקשרים בין מדדי לחץ, בעוד שמבחיני TF-IDF ו-N-grams סיפק תובנות על קוטביות הסנטימנטים. VADER-תפסו את העושר הסמנטי, וטכניקות הוויזואליזציה הציעו פרשנויות אינטואיטיביות, שהפכו נתונים מורכבים לנגישים יותר.

What did you find in your data through this representation?

English: Key findings include:

- A strong correlation between emotional overload and stress intensity, leading to the creation of an averaged metric.
- No significant difference in stress levels before and during protests (T-test $p\text{-value} > 0.5$).
- Frequent use of terms like "baby" and "pregnancy," and phrases like "I have a," highlighting emotional and medical discussions.
- Negative sentiment increased with higher stress levels, while positive sentiment remained unchanged.
- Posts with higher stress levels were significantly longer and used more punctuation, such as question marks.

Hebrew: ממצאים מרכזיים כוללים:

- קשר חזק בין עומס רגשי לעוצמת לחץ, שהוביל ליצירת משתנה ממוצע.
- לא נמצא הבדל משמעותי ברמות הלחץ לפני ובמהלך מחאות ($p\text{-value} > 0.5$).
- "I have a" ובביטויים כמו "pregnancy" ו-"baby" שימוש תדיר במילים כמו, המצביעים על דיונים רגשיים ורפואיים.
- סנטימנט שלילי עלה עם עליית רמות הלחץ, בעוד שסנטימנט חיובי נותר ללא שינוי.
- פוסטים עם רמות לחץ גבוהות היו ארוכים משמעותית והשתמשו יותר בפיסוק, כמו סימני שאלה.

Have you progressed with addressing your research question?

English: Partially. While the direct influence of pregnancy week on stress was not explicitly separated, our analyses revealed significant insights into how stress manifests in text patterns, sentiments, and word usage. These findings lay groundwork for future research to explore specific pregnancy weeks.

Hebrew: חלקית. בעוד שהשפעת שבוע ההיריון על הלחץ לא נבחנה במפורש, הניתוחים חשפו תובנות משמעותיות על איך הלחץ בא לידי ביטוי בדפוס טקסט, סנטימנטים ושימוש במילים. ממצאים אלו מהווים בסיס למחקרים עתידיים שיחקרו שבועות הריון ספציפיים.

Let me know if you need additional edits or elaborations!